Accumulating Transformations for Hierarchical Linear Regression HMM Adaptation

Field of Invention

[0001]   This invention relates to speech recognition and more particularly to adaptive speech recognition with hierarchical linear regression Hidden Markov Model (HMM) adaptation.

Background of Invention

[0002]   Hierarchical Linear Regression (HLR) (e.g. MLLR [See C. J. Leggetter and P.C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density HMMs," Computer, Speech and Language, 9(2):71-185, 1995]) is now a common technique to transform Hidden Markov Models(HMMs) for use in an acoustic environment different from the one in which the models are initially trained. The environments refer to speaker accent, speaker vocal tract, background noise, recording device, transmission channel,-etc. HLR improves word error rate (WER) substantially by reducing the mismatch between training and testing environments [See C.J. Leggetter cited above].

[0003]   Hierarchical Linear Regression (HLR) is a process that creates a set of transforms that can be used to adapt any subset of an initial set of Hidden Markov Models (HMMs) to a new acoustic environment. We refer to the new environment as the "target environment", and the adapted subset of HMM models as the "target models". The HLR adaptation process requires that some adaptation speech data from the new environment be collected, and converted into sequences of frames of vectors of speech parameters using well-known techniques. For example, to create a set of transforms to adapt an initial set of speaker-independent HMMs to a particular speaker who is using a particular microphone, adaptation speech data must be collected from the speaker and microphone, and then converted into frames of parameter vectors, such as the well-known cepstral vectors.

[0004]   There are two well known HLR methods for creating a set of transforms. In the first method, the adaptation speech data is aligned to states of the initial set of HMM models using well-known HMM Viterbi recognition alignment methods. A regression tree is formed which defines a hierarchical mapping from states of the initial HMM model set to linear transforms. Then the set of linear transforms is determined that adapts the initial HMM set so as to increase the likelihood of the adaptation speech data. While this method results in better